Crystal Structure Of ErbB2 And Uses Thereof

Field of the invention

The present invention relates generally to structural studies of ErbB2. More particularly, the present invention relates to the crystal structure of the ErbB2, in particular the crystal structure of an extracellular portion of ErbB2 and to methods of using the crystal and related structural information to screen for and design compounds that interact with or modulate ErbB2; or variants thereof.

Background to the invention

ErbB2 was discovered as an oncogene (neu) in a rat brain tumor (Schecter et al., 1984, Nature 312, 513-516). ErbB2/HER2 is closely related to the EGF receptor and is the most oncogenic member of the EGFR family. It is amplified and/or overexpressed in approximately 30% of human breast cancers and in many other types of human malignancies and this overexpression is correlated with poor clinical prognosis (see Mendelsohn and Baselga, 2000, Oncogene 19, 6550-6565; Yu and Hung, 2000, Oncogene 19, 6115-6121). Overexpression of ErbB2 enhances metastasis-related properties such as invasion, angiogenesis and increased survival of cancer cells, and confers increased resistance to various cancer therapies including chemotherapy and gamma-radiation (see Mendelsohn and Baselga, 2000; Yu and Hung, 2000). Some forms of breast cancer are now treated with antibodies that recognise ErbB2 and improvements in anti-ErbB2 therapies are likely to flow from a better understanding of its 3D structure and its mechanism of action.

Considerable resources have been directed to the identification of an ErbB2 ligand. No ligand has been found, however the search led to the discovery of ErbB4 and considerable improvements in our biological understanding of the EGF receptor family (Harari and Yarden, 2000, Oncogene 19, 6102-6114; Yarden and Sliwkowski, 2001, Nat. Rev. Mol. Cell. Biol. 2, 127-137). It now seems certain that ErbB2 has no ligand. Instead it acts as a second receptor sub-unit in three EGF receptor family heterodimers: ErbB1-ErbB2, ErbB3-ErbB2 and ErbB4-ErbB2 (Daly et al., 1997, Cancer Res. 57, 3804-3811; Sundaresan et al., 1998, Endocrinol. 139, 4756-4764). There is definitive evidence that the EGF receptor homodimer signals differently to the EGF receptor-

ErbB2 heterodimer. Unless ErbB2 carries an oncogenic mutation, as in c-neu, it signals only after activation of its heterodimer partner by EGF or other relevant ligand.

The human ErbB2 is a large (1234 residues), monomeric, modular glycoprotein with an extracellular domain, a single transmembrane region and an intracellular cytoplasmic tyrosine kinase, which is flanked by noncatalytic regulatory regions (Yamamoto et al., 1986, Nature 319, 230-234). The extracellular portion of human ErbB2 (residues 1-632), like the EGFR, consists of four sub-domains L1, CR1, L2 and CR2 (Bajaj et al., 1987, Biochim. Biophys. Acta 916, 220-226; Ward et al., 1995, Proteins: Struct. Funct. Genet. 22, 141-153) also referred to as domains I-IV (Lax et al., 1988, Mol. Cell. Biol. 8, 1970-1978).

Summary of the invention

We have determined the three dimensional structure of a truncated form (residues 1-509) of the ectodomain of the tyrosine kinase receptor ErbB2 at 2.5 Å resolution and compared it with the recently solved structures of the EGFR ectodomain with TGFα or EGF and the unliganded ErbB3 ectodomain. Lack of ligand binding by ErbB2 appears to be caused by amino acid differences in the L1 and L2 domains of ErbB2. Furthermore, ligands would not be able to bind to the observed conformation of ErbB2 here as kinks in the first Cys-rich region (CR1) lead to a closer juxtaposition of the L domains, occluding the region of ErbB2 that is analogous to the EGFR ligand binding site. The L1/L2 buried surface area and the degree of complementarity in the L domain interface implies that this "closed" form is biologically relevant.

Accordingly, in one aspect, the present invention provides a method for identifying a potential modulator compound for ErbB2 which method comprises:

(a) providing a three-dimensional structure of

(i) amino acids 1-509 of ErbB2 polypeptide having the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I; or

(ii) a subset of said amino acids having a corresponding subset of the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the

corresponding backbone atoms described by the atomic coordinates shown in Appendix I;

(b) providing the three-dimensional structure of a candidate compound;

(c) assessing the stereochemical complementarity between the three-dimensional structure of step (b) and a region of the three-dimensional structure of step (a); and

(d) selecting a compound on the basis of the stereochemical complementarity.

In a preferred embodiment, the method further comprises:

(e) synthesising or obtaining a candidate compound assessed in step (c) as possessing stereochemical complementarity with the three-dimensional structure of step (a);

(f) determining the ability of the candidate compound to interact with and/or modulate the activity of ErbB2.

In yet a further aspect the present invention provides a method for preparing a pharmaceutical composition for treating diseases associated with aberrant ErbB2 signalling, the method comprising:

(a) providing a three-dimensional structure of

(i) amino acids 1-509 of ErbB2 polypeptide having the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I; or

(ii) a subset of said amino acids having a corresponding subset of the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I;

(b) providing the three-dimensional structure of a candidate compound;

(c) assessing the stereochemical complementarity between the three-dimensional structure of step (b) and a region of the three-dimensional structure of step (a); and

(d) selecting a compound on the basis of the stereochemical complementarity;

(e) synthesising or obtaining a candidate compound assessed in step (c) as possessing stereochemical complementarity with the three-dimensional structure of step (a);

(f) determining the ability of the candidate compound to interact with and/or modulate the activity of ErbB2; and

(g) incorporating the compound into a pharmaceutical composition.

The method may be used for either targeted or broad screening. Targeted screening involves the design and synthesis of chemical compounds that are analogs of some active compounds or that can specifically act with the biological target under investigation. Broad screening involves the design and synthesis of a large array of maximally diverse chemical compounds, leading to diverse libraries that are tested against a variety of biological targets.

In a further aspect, the present invention provides a method of modulating ErbB2, the method comprising contacting the receptor with a compound that matches a region selected from at least one of the CR1 domain, the potential CR1 loop docking site between the L1, CR1 and L2 domains, the CR1-L2 hinge region, the regions of the L1 and L2 domains that contact each other in a closed conformation.

The compound may be a small molecule modulator. The term "small molecule" includes an organic compound either synthesized in the laboratory or found in nature. Typically, a small molecule is any organic molecule having a molecular weight of less than about 1500. Preferably the molecule has a molecular weight less that about 1000, more preferably less than about 500.

The term "ErbB2" as used herein includes wild-type ErbB2 and variants thereof including allelic variants and naturally occurring mutations and genetically engineered variants.

The present invention also provides a set of coordinates as shown in Appendix I, or a subset thereof, where said coordinates define a three dimensional structure of amino acids 1-509 of an ErbB2 polypeptide or a subset of said amino acids, or a set of coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I, or a subset thereof.

In a related aspect, the present invention provides a computer for producing a three-dimensional representation of a molecule or molecular complex, wherein the computer comprises:

(a) a machine-readable data storage medium comprising a data storage material encoded with machine-readable data, wherein the machine readable data comprises (i) the atomic coordinates of amino acids 1-509 of an ErbB2 polypeptide as shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I; or (ii) the atomic coordinates of a subset of said amino acids having a corresponding subset of the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I;

(b) a working memory for storing instructions for processing the machine-readable data;

(c) a central-processing unit coupled to the working memory and to the machine-readable data storage medium, for processing the machine-readable data into the three dimensional representation; and

(d) an output hardware coupled to the central processing unit, for receiving the three-dimensional representation.

Preferably, said subsets of amino acids are selected from the CR1 domain and the potential CR1 loop docking site between the L1, CR1 and L2 domains equivalent to that seen in the TGFα:sEGFR dimer complex (Garrett et al., 2002, Cell 110, 763-773), or the CR1-L2 hinge region or the regions of the L1 and L2 domains that contact each other in this closed conformation.

More preferably the subset of amino acids defines a homodimerisation or heterodimerisation surface with other EGF receptor family members. Preferred heterodimerisation surfaces include (i) the N-terminal end of the CR1 domain (residues 200-203, 210-213, 216-218, 225-230), (ii) the CR1 domain dimerisation loop (residues 247-268) and adjacent residues (residues 244-246, 285-289) and (iii) the C-terminal end of the CR1 domain (residues 294-319).

In a further preferred embodiment, the subset of amino acids comprises the following residues: Gln 36, Gln 60, Arg 82, Thr 84, Gln 85, Phe 237, Thr 269, Phe 270, Gly 271, Ala 272, Tyr 282, Thr 285, Gly 288, Ser 289, Cys 290, Thr 291, Leu 292, Val 293, Cys 294, Pro 295 and Cys 310.

The three-dimensional structure of ErbB2 may be used to develop models useful for drug design and *in silico* screening of candidate compounds that modulate ErbB2 activity. Other physicochemical characteristics may also be used in developing the model, e.g. bonding, electrostatics etc.

Generally the term "*in silico*" refers to the creation in a computer memory, i.e., on a silicon or other like chip. Stated otherwise "*in silico*" means "virtual." When used herein the term "*in silico*" is intended to refer to screening methods based on the use of computer models rather than *in vitro* or *in vivo* experiments.

By "modulate" we mean that the compound increases or decreases signal transduction via ErbB2. The phrase "decreases signal transduction" is intended to encompass partial or complete inhibition of signal transduction via ErbB2. The ability of a candidate compound to increase or decrease signal transduction via ErbB2 can be assessed by any one of the ErbB2 cell-based assays described herein.

The term "small molecule" includes a compound with a molecular weight of 1500 or less. Preferably, the small molecule has a molecular weight of less than 1000, particularly preferred is a molecule having a molecular weight of less than 500.

Accordingly, in yet a further aspect, the present invention provides a computer-based method of identifying a candidate modulator of ErbB2, which method comprises fitting the structure of

(i) amino acids 1-509 of an ErbB2 polypeptide having the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I; or

(ii) a subset of said amino acids having a corresponding subset of the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I;

to the structure of a candidate modulator molecule.

In a further related aspect, the present invention provides a computer-assisted method for identifying candidate compounds able to interact with ErbB2 and thereby modulate an activity mediated by the receptor, using a programmed computer comprising a processor, an input device, and an output device, which method comprises the steps of:

(a) entering into the programmed computer, through the input device, data comprising the atomic coordinates of amino acids 1-509 of ErbB2 as shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I, or a subset of said coordinates;

(b) generating, using computer methods, a set of atomic coordinates of a structure that possesses stereochemical complementarity to the atomic coordinates entered in step (a), thereby generating a criteria data set;

(c) comparing, using the processor, the criteria data set to a computer database of chemical structures;

(d) selecting from the database, using computer methods, chemical structures which are similar to a portion of said criteria data set; and

(e) outputting, to the output device, the selected chemical structures which are complementary to or similar to a portion of the criteria data set.

In another related aspect, the present invention provides a method for evaluating the ability of a chemical entity to interact with an ErbB2, said method comprising the steps of:

(a) providing a computer model of at least one region of ErbB2 using structure coordinates wherein the root mean square deviation between said structure coordinates and the structure coordinates of amino acids 1-509 of ErbB2 as set forth in Appendix I is not more than 1.5 Å;

(b) employing computational means to perform a fitting operation between the chemical entity and said computer model of the binding surface; and

(c) analysing the results of said fitting operation to quantify the association between the chemical entity and the binding surface model.

The model may be adaptive in a sense that it allows for slight surface changes to improve the fit between the candidate compound and the protein , e.g. by small movements in side chains or main chain.

Preferably, the region of ErbB2 is defined by the CR1 domain and the potential CR1 loop docking site between the L1, CR1 and L2 domains equivalent to that seen in the TGFα:sEGFR dimer complex (Garrett et al., 2002), or the CR1-L2 hinge region or the regions of the L1 and L2 domains that contact each other in this closed conformation and combinations thereof.

More preferably the region defines a heterodimerisation surface with other EGF receptor family members. Preferred heterodimerisation surfaces include (i) the N-terminal end of the CR1 domain (residues 200-203, 210-213, 216-218, 225-230), (ii) the CR1 domain dimerisation loop (residues 247-268) and adjacent residues (residues 244-246, 285-289) and (iii) the C-terminal end of the CR1 domain (residues 294-319).

In a further preferred embodiment, the region comprises the following amino acid residues: Gln 36, Gln 60, Arg 82, Thr 84, Gln 85, Phe 237, Thr 269, Phe 270, Gly 271, Ala 272, Tyr 282, Thr 285, Gly 288, Ser 289, Cys 290, Thr 291, Leu 292, Val 293, Cys 294, Pro 295 and Cys 310.

The ErbB2 crystal structure provided herein may also be used to model/solve the structure of a new crystal using molecular replacement. Accordingly, in a further aspect the present invention provides a method of using molecular replacement to obtain structural information about a molecule or a molecular complex of unknown structure, comprising the steps of:

(i) crystallising said molecule or molecular complex;

(ii) generating an X-ray diffraction pattern from said crystallized molecule or molecular complex;

(iii) applying at least a portion of the structure coordinates set forth in Appendix I, or structure coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the structure coordinates set forth in Appendix I, to the X-ray diffraction pattern to generate a three-dimensional electron density map of at least a portion of the molecule or molecular complex whose structure is unknown.

Preferably the molecule of unknown structure is ErbB2 or variant thereof.

In one embodiment, the molecular complex of unknown structure is a complex of ErbB2, or variant thereof, and a ligand or candidate ligand.

In another embodiment the molecular complex of unknown structure is a complex of ErbB2 and an EGF receptor. The molecular complex of unknown structure may also be a complex of ErbB2, an ErbB1 (EGF receptor), ErbB3 or ErbB4 receptor and a ligand or candidate ligand.

The screening methods of the fourth aspect of the invention may be used to identify compounds that modulate ErbB2 signalling. Such compounds may be used to treat disorders associated with ErbB2 dysfunction.

Accordingly, in a further aspect, the present invention provides a method for preventing or treating a disease associated with signaling by ErbB2 which method comprises administering to a subject in need thereof a compound identified by the screening methods of the invention.

The present invention also provides a pharmaceutical composition comprising a compound identified by the screening methods of the invention, which compound is able to bind to the extracellular domain of ErbB2 and modulate an activity of said receptor, as well as a method of preventing or treating a disease associated with signalling by ErbB2 which method comprises administering to a subject in need thereof a composition of the invention.

In yet a further aspect, the present invention provides a crystal of an ErbB2 polypeptide. In particular the present invention provides a crystal of an ErbB2 polypeptide having a space group $P2_12_12_1$ with unit cell dimensions of a=75.96 Å, b=82.24 Å, and c=110.06 Å with up to about 1% variation in any cell dimension. Preferably said ErbB2 polypeptide is a truncated soluble extracellular domain of the full-length ErbB2.

The present invention also provides a crystalline composition comprising a crystal of ErbB2.

In a further aspect, the invention provides a computer system for identifying one or more candidate modulators of ErbB2, the system containing data representing the structure of

(i) amino acids 1-509 of ErbB2 polypeptide having the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I; or

(ii) a subset of said amino acids having a corresponding subset of the atomic coordinates shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I.

The present invention further provides a computer readable media having recorded thereon data representing a model and/or the atomic coordinates of a ErbB2 crystal. Also provided is a computer readable media having recorded thereon coordinate data according to Appendix I, or a subset thereof, where said coordinate data define a three dimensional structure of amino acids 1-509 of ErbB2 polypeptide or a subset of said amino acids, or coordinate data having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinate according to Appendix I, or a subset thereof.

Particular diseases associated with signalling by ErbB2 include cancerous conditions such as cancer of the brain, head and neck, prostate, testicular, ovary, breast, cervix, lung, pancreas and colon; and melanoma, rhabdomyosarcoma, mesothelioma, squamous carcinomas of the skin and glioblastoma.

The information provided in Appendix I shows that there are a number of loop structures that line the ErbB2 dimerisation surface. It is envisaged that antibodies directed against these loop structures would interfere with the formation of heterodimers with other members of the EGF receptor family.

Accordingly, in a further aspect the present invention provides an antibody that binds to ErbB2, the antibody being directed against a structure defined by (i) ErbB2 amino acid residues 200-203, (ii) ErbB2 amino acid residues 210-213, (iii) ErbB2 amino acid residues 216-218, (iv) ErbB2 amino acid residues 225-230, (v) ErbB2 amino acid

residues 247-268 or a subset thereof; (vi) ErbB2 amino acid residues 244-246, (vii) ErbB2 amino acid residues 285-289, or (viii) ErbB2 amino acid residues 294-319 or a subset thereof.

In yet a further aspect the present invention provides an isolated conformationally constrained peptide or peptidomimetic consisting essentially of (i) ErbB2 amino acid residues 200-203, (ii) ErbB2 amino acid residues 210-213, (iii) ErbB2 amino acid residues 216-218, (iv) ErbB2 amino acid residues 225-230, (v) ErbB2 amino acid residues 247-268 or a subset thereof; (vi) ErbB2 amino acid residues 244-246, (vii) ErbB2 amino acid residues 285-289, or (viii) ErbB2 amino acid residues 294-319 or a subset thereof.

In yet a further aspect the present invention provides a computer-assisted method for identifying potential mimetics of ErbB2, using a programmed computer comprising a processor, a data storage system, an input device, and an output device, comprising the steps of:

(a) inputting into the programmed computer through said input device data comprising the atomic coordinates of amino acids 200-203, 210-213, 216-218, 225-230, 247-268, 244-246, 285-289, or 294-319 of ErbB2 as shown in Appendix I, or atomic coordinates having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I, thereby generating a criteria data set;

(b) comparing, using said processor, said criteria data set to a computer database of chemical structures stored in said computer data storage system;

(c) selecting from said database, using computer methods, chemical structures having a portion that is structurally similar to said criteria data set;

(d) outputting to said output device the selected chemical structures having a portion similar to said criteria data set.

In yet a further aspect the present invention provides a computer-assisted method for identifying potential mimetics of ErbB2, using a programmed computer comprising a processor, a data storage system, an input device, and an output device, comprising the steps of:

(a) inputting into the programmed computer through said input device data comprising the atomic coordinates of amino acids 200-203, 210-213, 216-218, 225-230, 247-268, 244-246, 285-289, or 294-319 of ErbB2 as shown in Appendix I, or atomic coordinates

having a root mean square deviation of backbone atoms of not more than 1.5Å when superimposed on the corresponding backbone atoms described by the atomic coordinates shown in Appendix I, thereby generating a criteria data set;
(b) constructing, using computer methods, a model of a chemical structure having a portion that is structurally similar to said criteria data set;
(c) outputting to said output device the constructed model.

In yet a further aspect the present invention provides a compound having a chemical structure selected using a method of the present invention, said compound being an ErbB2 mimetic. Preferably, the compound is a peptidomimetic that has fewer than 30 amino acids, more preferably fewer than 25 amino acids.

As will be readily understood by those skilled in this field the methods of the present invention provide a rational method for designing and selecting compounds including antibodies which interact with ErbB2. In the majority of cases these compounds will require further development in order to increase activity. Such further development is routine in this field and will be assisted by the structural information provided in this application. It is intended that in particular embodiments the methods of the present invention includes such further developmental steps.

It is also intended that embodiments of the present invention include manufacturing steps such as incorporating the compound into a pharmaceutical composition in the manufacture of a medicament.

Throughout this specification, preferred aspects and embodiments apply, as appropriate, separately, or in combination, to other aspects and embodiments, mutatis mutandis, whether or not explicitly stated as such.

Brief Description of the Figures

Figure 1. Structure-based sequence alignment of the human ErbB2 ectodomain with other members of the ErbB family.

(A) The receptor L1 and L2 domains plus the first module of the cys-rich regions, CR1 and CR2. Positions with conserved physicochemical properties of amino acids are boxed. Disulfide bond connections are shown are solid lines. Secondary structure

elements are indicated above and below the sequences as cylinders for α-helices and arrows for β-strands. Residues buried at L1/L2 interface are denoted by R. Sequence sources are: EGFR (Ullrich et al., 1984, Nature 309, 418-425), ErbB2 (Yamamoto et al., 1986); ErbB3 (Kraus et al., 1989, Proc Natl Acad Sci U S A. 86, 9193-9197; Plowman et al., 1990, Proc. Natl Acad. Sci. U S A. 87, 4905-4909); ErbB4 (Plowman et al., 1993, Proc. Natl. Acad. Sci. USA. 90, 1746-1750).

(B) Modules 2 to 8 of the ErbB family cys-rich region CR1 and modules 2 to 7 of CR2. Three types of disulfide bonded modules are indicated by bars below the sequences. The unfilled bars below parts of the cys-rich sequences indicate modules with 2 disulfide bonds (in a Cys 1-3 and 2-4 arrangement), the solid bars indicate modules which contain a single disulfide bond and have a β-finger motif, and the dashed bar indicates residues present in a disulfide-linked bend consisting of only five residues. Disulfide bonds are shown in solid lines and except for those that do not conform to the CR1 pattern which are indicated as dashed lines. The number in parentheses shows where amino acids have been omitted. Boxed residues and secondary structure elements are as in A.

Figure 2. Polypeptide fold for residues 1-509 of ErbB2 and its comparison with EGFR (1-501) as seen in the 2:2 complex with TGFα, and the full length ectodomain of ErbB3.

Figure 3. Percentage inhibition of thymidine incorporated in a cell line expressing erbB2 on EGFR-K721R (a kinase defective EGFR) + full length ErbB2 by compounds 39293, 94289, 19378 and 20697.

Detailed description of the invention

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g. in molecular biology, biochemistry, structural biology, and computational biology). Standard techniques are used for molecular and biochemical methods (see generally, Sambrook et al., Molecular Cloning: A Laboratory Manual, 3rd ed. (2001) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel et al., Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc. - and the full version

entitled Current Protocols in Molecular Biology, which are incorporated herein by reference) and chemical methods.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

ErbB2 crystals and crystal structures

The present invention provides a crystal comprising an ErbB2 polypeptide. Such crystals preferably are of the space group $P2_12_12_1$ with unit cell dimensions of a=75.96 Å, b=82.24 Å, and c=110.06 Å.

As used herein, the term "crystal" means a structure (such as a three dimensional (3D) solid aggregate) in which the plane faces intersect at definite angles and in which there is a regular structure (such as internal structure) of the constituent chemical species. The term "crystal" refers in particular to a solid physical crystal form such as an experimentally prepared crystal.

Crystals according to the invention may be prepared using full-length ErbB2 polypeptides. However, preferably the extracellular domain is employed in isolation. Thus, preferably the ErbB2 polypeptide is a truncated polypeptide containing the extracellular domain and lacking the transmembrane domain and the intracellular tyrosine kinase domain. Typically, the extracellular domain comprises residues 1 to 632 (mature receptor numbering) of human ErbB2, or the equivalent thereof, or a truncated version thereof, preferably comprising amino acids 1 to 509, or the equivalent residues in other ErbB2 polypeptides.

In a preferred embodiment the ErbB2 polypeptide is human ErbB2 (Accession No. A24571 – mature protein begins at residue 22). However, the ErbB2 polypeptide may also be obtained from other species, such as other mammalian species.

Crystals may be constructed with wild-type ErbB2 polypeptide sequences or variants thereof, including allelic variants and naturally occurring mutations as well as genetically engineered variants. Typically, variants have at least 95 or 98% sequence identity with a corresponding wild-type ErbB2 polypeptide.

Optionally, the crystal of ErbB2 may comprise one or more molecules which bind to ErbB2, or otherwise soaked into the crystal or cocrystallise with ErbB2. Such molecules include ligands or small molecules, which may be candidate pharmaceutical agents intended to modulate the interaction between ErbB2 and its biological targets or dimer partners, such as other members of the EGF receptor family. The crystal of ErbB2 may also be a molecular complex with other receptors of the EGF receptor family such as ErbB1 (the EGF receptor), ErbB3 or ErbB4. The complex may also comprise additional molecules such as the ligands to these receptors.

The production of ErbB2 crystals is described below.

In a preferred embodiment, an ErbB2 crystal of the invention has the atomic coordinates set forth in Appendix I. It will be understood by those skilled in the art that atomic coordinates may be varied, without affecting significantly the accuracy of models derived therefrom; thus, although the invention provides a very precise definition of a preferred atomic structure, it will be understood that minor variations are envisaged and the claims are intended to encompass such variations. Preferred are variants in which the r.m.s. deviation of the x, y and z co-ordinates for all backbone atoms other than hydrogen is less than 1.5 Å (preferably less than 1 Å, 0.7 Å or less than 0. 3 Å) compared with the coordinates given in Appendix I.

In a highly preferred embodiment, the crystal has the atomic coordinates as shown in Appendix I.

As used herein, the term "atomic co-ordinates" refer to a set of values which define the position of one or more atoms with reference to a system of axes.

The present invention also provides a crystal structure of an ErbB2 polypeptide, in particular a crystal structure of the extracellular domain of an ErbB2 polypeptide, or a region thereof.

The atomic coordinates obtained experimentally for amino acids 1 to 509 (mature receptor numbering) of human ErbB2 are shown in Appendix I. However, a person skilled in the art will appreciate that a set of atomic coordinates determined by X-ray crystallography is not without standard error. Accordingly, any set of structure coordinates for an ErbB2 polypeptide that has a root mean square deviation of protein backbone atoms of less than 0.75 Å when superimposed (using backbone atoms) on the atomic coordinates listed in Appendix I shall be considered identical.

The present invention also comprises the atomic coordinates of an ErbB2 polypeptide that substantially conform to the atomic coordinates listed in Appendix I.

A structure that "substantially conforms" to a given set of atomic coordinates is a structure wherein at least about 50% of such structure has an average root-mean-square deviation (RMSD) of less than about 1.5 Å for the backbone atoms in secondary structure elements in each domain, and more preferably, less than about 1.3 Å for the backbone atoms in secondary structure elements in each domain, and, in increasing preference, less than about 1.0 Å, less than about 0.7 Å, less than about 0.5 Å, and most preferably, less than about 0.3 Å for the backbone atoms in secondary structure elements in each domain.

In a more preferred embodiment, a structure that substantially conforms to a given set of atomic coordinates is a structure wherein at least about 75% of such structure has the recited average root-mean-square deviation (RMSD) value, and more preferably, at least about 90% of such structure has the recited average root-mean-square deviation (RMSD) value, and most preferably, about 100% of such structure has the recited average root-mean-square deviation (RMSD) value.

In an even more preferred embodiment, the above definition of "substantially conforms" can be extended to include atoms of amino acid side chains. As used herein, the phrase "common amino acid side chains" refers to amino acid side chains that are common to both the structure which substantially conforms to a given set of atomic coordinates and the structure that is actually represented by such atomic coordinates.

The present invention also provides subsets of said atomic coordinates listed in Appendix I and subsets that conform substantially thereto. Preferred subsets define one or more regions of the human ErbB2 extracellular domain selected from the CR1

domain and the potential CR1 loop docking site between the L1, CR1 and L2 domains equivalent to that seen in the TGFα:sEGFR dimer complex (Garrett et al., 2002), or the CR1-L2 hinge region or the regions of the L1 and L2 domains that contact each other in this closed conformation. A particularly preferred subset defines the heterodimerisation surface of ErbB2 with other members of the EGF receptor family, such as ErbB1, ErbB3 and/or ErbB4.

It will be appreciated that a set of structure coordinates for a polypeptide is a relative set of points that define a shape in three dimensions. Thus, it is possible that an entirely different set of coordinates could define a similar or identical shape. Moreover, slight variations in the individual coordinates will have little effect on overall shape.

The variations in coordinates may be generated due to mathematical manipulations of the structure coordinates. For example, the structure coordinates set forth in Appendix I could be manipulated by crystallographic permutations of the structure coordinates, fractionalisation of the structure coordinates, integer additions or subtractions to sets of the structure coordinates, inversion of the structure coordinates, or any combination thereof.

Alternatively, modification in the crystal structure due to mutations, additions, substitutions, and/or deletions of amino acids, or other changes in any of the components that make up the crystal could also account for variations in structure coordinates.

Various computational analyses are used to determine whether a molecular complex or a portion thereof is sufficiently similar to all or parts of the structure of the extracellular domain of ErbB2 described above. Such analyses may be carried out in current software applications, such as the Molecular Similarity program of QUANTA (Molecular Simulations Inc., San Diego, CA) version 4.1.

The Molecular Similarity program permits comparisons between different structures, different conformations of the same structure, and different parts of the same structure.

Comparisons typically involve calculation of the optimum translations and rotations required such that the root mean square difference of the fit over the specified pairs of equivalent atoms is an absolute minimum. This number is given in angstroms.

Accordingly, structural coordinates of an ErbB2 within the scope of the present invention include structural coordinates related to the atomic coordinates listed in Appendix I by whole body translations and/or rotations. Accordingly, r.m.s deviations listed above assume that at least the backbone atoms of the structures are optimally superimposed which may require translation and/or rotation to achieve the required optimal fit from which to calculate the r.m.s.d.

A three dimensional structure of an ErbB2 protein or region thereof which substantially conforms to a specified set of atomic coordinates can be modeled by a suitable modeling computer program such as MODELER (Sali and Blundell, 1993, J. Mol. Biol., vol. 234:779-815), as implemented in the Insight II Homology software package (Insight II (97.0), MSI, San Diego)), using information, for example, derived from the following data: (1) the amino acid sequence of the human ErbB2 protein; (2) the amino acid sequence of the related portion(s) of the protein represented by the specified set of atomic coordinates having a three dimensional configuration; and, (3) the atomic coordinates of the specified three dimensional configuration. A three dimensional structure of an ErbB2 protein which substantially conforms to a specified set of atomic coordinates can also be calculated by a method such as molecular replacement, which is described in detail below.

Structure coordinates/atomic coordinates are typically loaded onto a machine readable-medium for subsequent computational manipulation. Thus models and/or atomic coordinates are advantageously stored on machine-readable media, such as magnetic or optical media and random-access or read-only memory, including tapes, diskettes, hard disks, CD-ROMs and DVDs, flash memory cards or chips, servers and the internet. The machine is typically a computer.

The structure coordinates/atomic coordinates may be used in a computer to generate a representation, e.g. an image, of the three-dimensional structure of the ErbB2 crystal which can be displayed by the computer and/or represented in an electronic file.

The structure coordinates/atomic coordinates and models derived therefrom may also be used for a variety of purposes such as drug discovery and X-ray crystallographic analysis of other protein crystals.

Design/selection of chemical entities that bind ErbB2

Using a variety of known modelling techniques, the crystal structure of the present invention can be used to produce a model for at least part of ErbB2 .

As used herein, the term "modelling" includes the quantitative and qualitative analysis of molecular structure and/or function based on atomic structural information and interaction models. The term "modelling" includes conventional numeric-based molecular dynamic and energy minimisation models, interactive computer graphic models, modified molecular mechanics models, distance geometry and other structure-based constraint models.

Molecular modelling techniques can be applied to the atomic coordinates of the ErbB2 to derive a range of 3D models and to investigate the structure of binding sites, such as the binding sites of monoclonal antibodies and inhibitory peptides.

These techniques may also be used to screen for or design small and large chemical entities which are capable of binding ErbB2 and modulating the ability of ErbB2 to interact with extracellular biological targets, such as other members of the EGF receptor family e.g. which modulate the ability of ErbB2 to heterodimerise. The screen may employ a solid 3D screening system or a computational screening system.

Such modelling methods are to design or select chemical entities that possess stereochemical complementary to particular regions of ErbB2.

By "stereochemical complementarity" we mean that the compound or a portion thereof makes a sufficient number of energetically favourable contacts with the receptor as to have a net reduction of free energy on binding to the receptor.

Such stereochemical complementarity is characteristic of a molecule that matches intra-site surface residues lining the groove of the receptor site as enumerated by the coordinates set out in Appendix I. By "match" we mean that the identified portions interact with the surface residues, for example, via hydrogen bonding or by non-covalent Van der Waals and Coulomb interactions (with surface or residue) which promote desolvation of the molecule within the site, in such a way that retention of the molecule within the groove is favoured energetically.

It is preferred that the stereochemical complementarity is such that the compound has a K_d for the receptor site of less than 10^{-4}M, more preferably less than 10^{-5}M and more preferably 10^{-6}M. In a most preferred embodiment, the K_d value is less than 10^{-8}M and more preferably less than 10^{-9}M.

Chemical entities which are complementary to the shape and electrostatics or chemistry of the receptor site characterised by amino acids positioned at atomic coordinates set out in Appendix I will be able to bind to the receptor, and when the binding is sufficiently strong, substantially prohibit the interaction of the ErbB2 with biological target molecules such as other EGF receptors.

It will be appreciated that it is not necessary that the complementarity between chemical entities and the receptor site extend over all residues lining the groove in order to inhibit binding of a molecule or complex that naturally interacts with ErbB2.

A number of methods may be used to identify chemical entities possessing stereo-complementarity to a region of the extracellular domain of ErbB2. For instance, the process may begin by visual inspection of potential binding sites, for example, the binding sites for anti- ErbB2 antibodies, on the computer screen based on the ErbB2 coordinates in Appendix I generated from the machine-readable storage medium. Alternatively, selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within an individual binding site of ErbB2, as defined *supra*. Modelling software that is well known and available in the art may be used (Guida, W. C. (1994). "Software For Structure-Based Drug Design." Curr. Opin. Struct. Biology 4: 777-781). These include QUANTA and InsightII [Molecular Simulations, Inc., San Diego, Calif., a division of Pharmacopiea, Inc., Princeton, N.J., 1992], SYBYL [Molecular Modeling Software, Tripos Associates, Inc., St. Louis, Mo., 1992], This modelling step may be followed by energy minimization with standard molecular mechanics force fields such as AMBER [S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, and P. Weiner, J. Am. Chem. Soc., vol. 106, pp. 765-784 (1984)], and CHARMM [B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S Swaminathan, and M. Karplus, J. Comp. Chem. vol. 4, pp. 187-217 (1983)]. In addition, there are a number of more specialized computer programs to assist in the process of selecting the binding moieties of this invention.

Specialised computer programs may also assist in the process of selecting fragments or chemical entities. These include, inter alia:

1. GRID (Goodford, P. J.,"A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem., 28, pp. 849-857 (1985)). GRID is available from Oxford University, Oxford, UK.

2. MCSS (Miranker, A. and M. Karplus, "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. "Proteins: Structure, Function and Genetics, 11, pp. 29-34 (1991)). MCSS is available from Molecular Simulations, Burlington, MA.

3. AUTODOCK (Goodsell, D. S. and A. J. Olsen, "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8, pp. 195-202 (1990)). AUTODOCK is available from Scripps Research Institute, La Jolla, CA.

4. DOCK (Kuntz, I. D. et al.,"A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161, pp. 269-288 (1982)). DOCK is available from University of California, San Francisco, CA.

Once suitable chemical entities or fragments have been selected, they can be assembled into a single compound. In one embodiment, assembly may proceed by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of ErbB2. This is followed by manual model building using software such as Quanta or Sybyl. Alternatively, fragments may be joined to additional atoms using standard chemical geometry.

The above-described evaluation process for chemical entities may be performed in a similar fashion for chemical compounds.

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include:

1. CAVEAT (Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules". In "Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc., 78, pp. 182-196 (1989)). CAVEAT is available from the University of California, Berkeley, CA.

2. 3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, CA). This area is reviewed in Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35, pp. 2145-2154 (1992)).

3. HOOK (available from Molecular Simulations, Burlington, MA).

Other molecular modeling techniques may also be employed in accordance with this invention. See, e. g., Cohen et al., "Molecular Modeling Software and Methods for Medicinal Chemistry", J. Med. Chem., 33, pp. 883-894 (1990). See also Navia and Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2, pp. 202-210 (1992).

There are two preferred approaches to designing a molecule, according to the present invention, that complement the stereochemistry of ErbB2. The first approach is to in silico directly dock molecules from a three-dimensional structural database, to the receptor site, using mostly, but not exclusively, geometric criteria to assess the goodness-of-fit of a particular molecule to the site. In this approach, the number of internal degrees of freedom (and the corresponding local minima in the molecular conformation space) is reduced by considering only the geometric (hard-sphere) interactions of two rigid bodies, where one body (the active site) contains "pockets" or "grooves" that form binding sites for the second body (the complementing molecule).

This approach is illustrated by Kuntz et al., 1982, J. Mol. Biol. 161: 269, and Ewing et al., 2001, J. Comput-Aid. Mol. Design 15: 411, the contents of which are hereby incorporated by reference, whose algorithm for ligand design is implemented in a commercial software package, DOCK version 4.0, distributed by the Regents of the University of California and further described in a document, provided by the distributor, which is entitled "Overview of the DOCK program suite" the contents of which are hereby incorporated by reference. Pursuant to the Kuntz algorithm, the shape of the cavity represented by a site on ErbB2 is defined as a series of overlapping spheres of different radii. One or more extant databases of crystallographic data, such

as the Cambridge Structural Database System maintained by Cambridge University (University Chemical Laboratory, Lensfield Road, Cambridge, U.K.), the Protein Data Bank maintained by the Research Collaboratory for Structural Bioinformatics (Rutgers University, N.J., U.S.A.), LeadQuest (Tripos Associates, Inc., St. Louis, MO), Available Chemicals Directory (Molecular Design Ltd., San Leandro, CA), and the NCI database (National Cancer Institute, U.S.A) is then searched for molecules which approximate the shape thus defined.

Molecules identified on the basis of geometric parameters, can then be modified to satisfy criteria associated with chemical complementarity, such as hydrogen bonding, ionic interactions and Van der Waals interactions. Different scoring functions can be employed to rank and select the best molecule from a database. See for example Bohm and Stahl, 1999, M. Med. Chem. Res. 9: 445. The software package FlexX, marketed by Tripos Associates, Inc. (St. Louis, MO) is another program that can be used in this direct docking approach (see Rarey, M. et al., J. Mol. Biol. 1996, 261: 470).

The second preferred approach entails an assessment of the interaction of respective chemical groups ("probes") with the active site at sample positions within and around the site, resulting in an array of energy values from which three-dimensional contour surfaces at selected energy levels can be generated. The chemical-probe approach to ligand design is described, for example, by Goodford, 1985, J. Med. Chem. 28:849, the contents of which are hereby incorporated by reference, and is implemented in several commercial software packages, such as GRID (product of Molecular Discovery Ltd., West Way House, Elms Parade, Oxford OX2 9LL, U.K.).

Pursuant to this approach, the chemical prerequisites for a site-complementing molecule are identified at the outset, by probing the active site with different chemical probes, e.g., water, a methyl group, an amine nitrogen, a carboxyl oxygen, or a hydroxyl. Favoured sites for interaction between the active site and each probe are thus determined, and from the resulting three-dimensional pattern of such sites a putative complementary molecule can be generated. This may be done either by programs that can search three-dimensional databases to identify molecules incorporating desired pharmacophore patterns or by programs which using the favoured sites and probes as input to perform de novo design. Suitable programs for determining and designing pharmacophores include CATALYST (including HypoGen or HipHop) (Molecular

Simulations, Inc), and CERIUS2, DISCO (Abbott Laboratories, Abbott Park, IL) and ChemDBS-3D (Chemical Design Ltd., Oxford, U.K.).

The pharmacophore can be used to screen in silico compound libraries/ three-dimensional databases, using a program such as CATALYST (Molecular Simulations, Inc); MACCS-3D and ISIS/3D (Molecular Design Ltd., San Leandro, CA), ChemDBS-3D (Chemical Design Ltd., Oxford, U.K.), and Sybyl/3DB Unity (Tripos Associates, Inc., St. Louis, MO).

Databases of chemical structures are available from a number of sources including Cambridge Crystallographic Data Centre (Cambridge, U.K.), Molecular Design, Ltd., (San Leandro, CA), Tripos Associates, Inc. (St. Louis, MO), Chemical Abstracts Service (Columbus, OH), the Available Chemical Directory (MDL Inc), the Derwent World Drug Index (WDI), BioByteMasterFile, the National Cancer Institute database (NCI), and the Maybridge catalogue.

De novo design programs include LUDI (Biosym Technologies Inc., San Diego, CA), Leapfrog (Tripos Associates, Inc.), Aladdin (Daylight Chemical Information Systems, Irvine, CA), and LigBuilder (Peking University, China).

Once an entity or compound has been designed or selected by the above methods, the efficiency with which that entity or compound may bind to ErbB2 can be tested and optimised by computational evaluation. For example, a compound that has been designed or selected to function as an ErbB2 binding compound must also preferably traverse a volume not overlapping that occupied by the binding site when it is bound to the native ErbB2. An effective ErbB2 binding compound must preferably demonstrate a relatively small difference in energy between its bound and free states (i. e., a small deformation energy of binding). Thus, the most efficient ErbB2 binding compound should preferably be designed with a deformation energy of binding of not greater than about 10 kcal/mole, preferably, not greater than 7 kcal/mole. ErbB2 binding compounds may interact with ErbB2 in more than one conformation that is similar in overall binding energy. In those cases, the deformation energy of binding is taken to be the difference between the energy of the free compound and the average energy of the conformations observed when the compound binds to the protein.

A compound designed or selected as binding to ErbB2 may be further computationally optimised so that in its bound state it would preferably lack repulsive electrostatic interaction with the target protein.

Such non-complementary (e.g., electrostatic) interactions include repulsive charge-charge, dipole-dipole and charge-dipole interactions. Specifically, the sum of all electrostatic interactions between the compound and the protein when the compound is bound to ErbB2, preferably make a neutral or favourable contribution to the enthalpy of binding.

Once an ErbB2-binding compound has been optimally selected or designed, as described above, substitutions may then be made in some of its atoms or side groups to improve or modify its binding properties. Generally, initial substitutions are conservative, i. e., the replacement group will have approximately the same size, shape, hydrophobicity and charge as the original group. It should, of course, be understood that components known in the art to alter conformation should be avoided. Such substituted chemical compounds may then be analysed for efficiency of fit to ErbB2 by the same computer methods described in detail above.

Specific computer software is available in the art to evaluate compound deformation energy and electrostatic interaction. Examples of programs designed for such uses include: Gaussian 92, revision C (Frisch, Gaussian, Inc., Pittsburgh, PA) ; AMBER, version 4.0 (Kollman, University of California at San Francisco); QUANTA/CHARMM (Molecular Simulations, Inc., Burlington, MA); and Insight II/Discover (Biosysm Technologies Inc., San Diego, CA).

The screening/design methods may be implemented in hardware or software, or a combination of both. However, preferably, the methods are implemented in computer programs executing on programmable computers each comprising a processor, a data storage system (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. Program code is applied to input data to perform the functions described above and generate output information. The output information is applied to one or more output devices, in known fashion. The computer may be, for example, a personal computer, microcomputer, or workstation of conventional design.

Each program is preferably implemented in a high level procedural or object-oriented programming language to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. In any case, the language may be compiled or interpreted language.

Each such computer program is preferably stored on a storage medium or device (e.g., ROM or magnetic diskette) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform the functions described herein.

Compounds identified by, or designed by the methods of the invention can be synthetic or naturally occurring, preferably synthetic. In one embodiment, a synthetic compound selected or designed by the methods of the invention preferably has a molecular weight equal to or less than about 5000 or 1000 daltons. A compound selected or designed by methods of this invention is preferably soluble under physiological conditions.

Confirmation of binding and biological activity

Compounds selected or designed in accordance with the in silico methods of the invention may be subjected to further confirmation of binding to ErbB2 by cocrystallization of the compound with ErbB2 and structural determination, as described herein.

Compounds designed or selected according to the methods of the present invention are preferably assessed by a number of *in vitro* and *in vivo* assays of ErbB2 function to confirm their ability to interact with and modulate ErbB2 activity. For example, compounds may be tested for their ability to bind to ErbB2 and/or for their ability to modulate e.g. disrupt, heterodimerisation of ErbB2 to other members of the EGF receptor family such as ErbB1, ErbB3 or ErbB4.

Suitable assays include *in vitro* binding assays and ErbB2-dependent proliferation assays, such as described by Deb et al., 2001, J Biol Chem 276:15554-15560 or Berezov et al., 2001, J. Med. Chem. 44: 2565-2574.

Particular examples of suitable assays are described below.

Inhibition of heterodimer formation between ErbB2 and other ErbB family members

Rationale: While ErB2 is a major oncogenic therapeutic target in its own right, it is now clear that part of the tumour-promoting activity associated with ErbB2 often depends on ligand-induced heterodimer formation with other ErbB family members. There is no known ligand for ErbB2, however ligand binding to other ErbB family members (ErbB1, ErbB3 and ErbB4) causes their heterodimerization with ErbB2. Thus reagents that block this association, for example the ErbB2-specific antibody 2C4, inhibit ligand-stimulated growth *in vitro* and tumour xenograft *in vivo* (Agus, D.B. et.al. Cancer Cell 2:127-137). Heterodimerization results in cross-phosphorylation by the ErbB2 kinase of the dimerization partner. In particular, ErbB3 mediated signalling requires heterodimer formation as this particular ErbB family member lacks a functional kinase. Thus, while it is not possible to directly ligand-activate the ErbB2 kinase, it is possible to monitor its activity in cells co-expressing ErbB2 with one or more members of the EGFR family by adding ligands specific to the heterodimerization partners.

Methods: a number of readouts can be used to assess the efficacy, and specificity, of ErbB2 compounds/antibodies in cell-based assays of ligand-induced heterodimer formation. Activity can be assessed by one or more of the following:

(i) Inhibition of ligand-induced heterodimerisation of ErbB2 with other ErbB family members in a target cell line, for example MCF-7 breast cancer cells. Immunoprecipitation of ErbB2 complexes from cell lysates can be performed with a receptor-specific antibody, and the absence/presence of other ErbB receptors and their biologically relevant ligands within the complex can be analysed following electrophoresis/Western transfer by probing with antibodies to other ErbB receptors.

(ii) Inhibition of the activation of signalling pathways by ligand-activated heterodimers. Association with ErbB2 appears critical for other members of the ErbB

family of receptors to elicit maximal cellular response following ligand binding. In the case of the kinase-defective ErbB3, ErbB2 provides a functional tyrosine kinase domain to enable signalling to occur following binding of growth factor ligands. Thus, cells co-expressing ErbB2 and ErbB3 can be treated with ligand, for example heregulin, in the absence and presence of inhibitor and the effect on ErbB3 tyrosine phosphorylation monitored by a number of ways including immunoprecipitation of ErbB3 from treated cell lysates and subsequent Western blotting using anti-phosphotyrosine antibodies (see Agus op. cit. for details). Alternatively, a high-throughput assay can be developed by trapping ErbB3 from solubilized lysates onto the wells of a 96-well plate coated with an anti-ErbB3 receptor antibody, and the level of tyrosine phosphorylation measured using, for example, europium-labelled anti-phosphotyrosine antibodies, as embodied by Waddleton, D. et.al. Anal. Biochem. 309:150-157, 2002.

In a broader extension of this approach, effector molecules known to be activated downstream of activated receptor heterodimers, such as mitogen-activated protein kinases (MAPK) and Akt, may be analysed directly, by immunoprecipitation from treated lysates and blotting with antibodies that detect the activated forms of these proteins, or by analysing the ability of these proteins to modify/activate specific substrates.

(iii) Inhibition of ligand-induced cellular proliferation. A variety of cell lines are known to co-express combinations of ErbB receptors, for example many breast and prostate cancer cell lines. Assays may be performed in 24/48/96-well formats with the readout based around DNA synthesis (tritiated thymidine incorporation), increase in cell number (crystal violet staining) etc. However, co-expression of ErbB1 or ErbB4 in such cell lines will mean that it is difficult to determine whether ErbB1 or ErbB4 homodimer signalling is responsible for the proliferative response to ligand.

A new, semi-automated assay system to monitor ErbB2 signalling activity that may be used to confirm the ability of candidate compounds to interact with and modulate ErbB2 activity has been developed. This assay exploits the heterodimerization characteristic of the ErbB family of receptor. We have created a BaF/3 cell line, which normally does not express any members of the ErbB family and is IL-3 dependent, that co-expresses wild-type ErbB2 and a kinase defective (but ligand responsive) ErbB-1 mutant (EGFR-K721R). Upon exposure of the cells to EGF (or other ErbB1 ligand),

heterodimer formation occurs leading to phosphorylation of the kinase-defective ErbB1 by the ErbB2 kinase, initiation of the signal transduction pathways downstream of the receptors and ultimately to DNA synthesis. In this experimental system signalling is strictly ligand-dependent but is entirely mediated by the ErbB2 kinase, providing a specific and sensitive assay for inhibitors of ErbB2 heterodimerization. Non-specific toxicity of the test samples is assessed in parallel by testing the cells' responsiveness to IL-3 in the absence of EGF.

Method: BaF/3 cells co-expressing EGFR-K721R and full length wild type ErbB2 are routinely grown in RPMI/10%FCS containing IL-3. Before assay, cells are washed three times to remove residual IL-3 and resuspended in RPMI 1640 + 10% FCS. Cells are seeded into 96 well plates using a Biomek 2000 (Beckman) at $2x10^4$ cells per 200µl and incubated for 4 hours at 37°C in 10% CO_2. Putative ErbB2 inhibitors are added to the first titration point and titrated in two-fold dilutions across the 96 well plate in duplicate with or without a constant amount of EGF (1nM) or IL-3 (1µl). ^{3}H-Thymidine (0.5µCi/well) is added and the plates incubated for 20 hours at 37°C in 5% CO_2. At the end of the incubation the cells are lysed in 0.5M NaOH at room temperature for 30 minutes then harvested onto nitrocellulose filter mats using an automatic harvester (Tomtec, Connecticut, USA). The mats are dried, placed in a plastic counting bag and scintillant (10ml) added. Incorporated 3H-Thymidine is determined using a beta counter (1205 Betaplate, Wallac, Finland).

(iv) Inhibition of growth in soft-agar. This is the benchmark-type assay undertaken to assess anti-tumour activity prior to xenograft studies in animals. Cells are seeded into liquid soft agar cultures, the agar allowed to set, and the appearance of cell colonies monitored over the next 14-21 days. The appearance of colonies in semi-solid media is known as anchorage-independent growth, and is characteristic of the tumour phenotype. Cultures of tumour cell lines can be set up in the presence of both ligand and candidate antagonists of receptor heterodimerisation, and colony growth monitored.

(v) Ability of candidate compounds to block *in vivo* growth of tumour xenografts of human tumour cell lines whose tumorigenic phenotype is known to be at least partly dependent on ligand activation of ErbB2 heterodimer cell signalling e.g. MCF7 breast cancer cells, LNCap prostate cancer cells etc. This can be assessed in

immunocompromised mice either alone or in combination with an appropriate cytotoxic agent for the cell line in question.

Modulation of ligand-receptor interaction

Rationale & method: ErbB2 has no identified ligand of its own, yet in association with other ErbB family members can markedly influence the interaction of its heterodimer partner with ligand.

(i) Heterodimers of ErbB2/3, either on the cell-surface or generated as recombinant fusion proteins using an immunoglobulin Fc domain, bind heregulin with 2-3 orders of magnitude higher affinity than the equivalent ErbB3 homodimers (Jones, J.T. et. al. FEBS Lett. 447:227-231, 1999). Similarly, ErbB4 homodimers do not bind EGF, whereas ErbB2/4 heterodimers do (Jones op.cit.). The heterodimer antagonist antibody 2C4 blocks heregulin binding to cell-surface and Fc fusion heterodimers very efficiently, possibly as a result of steric hindrance through the ligand-binding site, although this remains to be established. This observation suggests that candidate inhibitors of heterodimer association, in particular the ErbB2 CR1 loop-specific antibodies can be tested for activity in this manner. Hence, it is possible to assay in a 96-well format the ability of lead entities (which may or may not be antibodies) to block the binding of tagged ligand, for example europium-labelled EGF, to immobilised ErbB2 hetrodimer combinations, in one example ErbB2/4 Fc fusion proteins, using time-resolved fluorescence as a readout.

(ii) Berezov, A. et. al. J. Biol. Chem. 277: 28330-28339 (2002) describe a screen using the BIAcore whereby small ErbB2 peptide mimetics are used to inhibit heterodimer formation between immobilised ErbB1, 2 or 3 ectodomains and a solution containing ErbB3 ectodomain and ligand (heregulin). The peptides are derived from the C-terminal region of the second cysteine-rich domain of ErbB2.

Molecular replacement/binding

The structure coordinates of ErbB2, such as those set forth in Appendix I, can also be used for determining at least a portion of the three-dimensional structure of a molecular complex which contains at least some structural features similar to at least a portion of ErbB2. In particular, structural information about another crystallised molecular

complex may be obtained. This may be achieved by any of a number of well-known techniques, including molecular replacement.

Methods of molecular replacement are generally known by those of skill in the art (generally described in Brunger, Meth. Enzym., vol. 276, pp. 558-580, 1997; Navaza and Saludjian, Meth. Enzym., vol. 276, pp. 581-594, 1997; Tong and Rossmann, Meth. Enzym., vol. 276, pp. 594-611, 1997; Bentley, Meth. Enzym., vol. 276, pp. 611-619, 1997); Lattman, "Use of the Rotation and Translation Functions", in Meth. Enzymol., 115, pp. 55-77 (1985); and Rossmann, ed., "The Molecular Replacement Method", Int. Sci. Rev. Ser., No. 13, Gordon & Breach, New York (1972)).

Generally, X-ray diffraction data are collected from the crystal of a crystallised target structure. The X-ray diffraction data is transformed to calculate a Patterson function. The Patterson function of the crystallised target structure is compared with a Patterson function calculated from a known structure (referred to herein as a search structure). The Patterson function of the crystallised target structure is rotated on the search structure Patterson function to determine the correct orientation of the crystallised target structure in the crystal. The translation function is then calculated to determine the location of the target structure with respect to the crystal axes. Once the crystallised target structure has been correctly positioned in the unit cell, initial phases for the experimental data can be calculated. These phases are necessary for calculation of an electron density map from which structural differences can be observed and for refinement of the structure. Preferably, the structural features (e.g., amino acid sequence, conserved di-sulphide bonds, and beta-strands or beta-sheets) of the search molecule are related to the crystallised target structure.

The electron density map can, in turn, be subjected to any well-known model building and structure refinement techniques to provide a final, accurate structure of the unknown crystallised molecular complex (eg see Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A 47,110–119; Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921).

Obtaining accurate values for the phases, by methods other than molecular replacement, is a time-consuming process that involves iterative cycles of approximations and refinements and greatly hinders the solution of crystal structures. However, when the crystal structure of a protein containing at least a homologous portion has been solved, the phases from the known structure provide a satisfactory estimate of the phases for the unknown structure.

By using molecular replacement, all or part of the structure coordinates of ErbB2 provided herein (and set forth in Appendix) can be used to determine the structure of a crystallised molecular complex whose structure is unknown more rapidly and efficiently than attempting to determine such information ab initio. This method is especially useful in determining the structure of ErbB2 mutants and homologues.

The structure of any portion of any crystallised molecular complex that is sufficiently homologous to any portion of the extracellular domain of ErbB2 can be solved by this method.

Such structure coordinates are also particularly useful to solve the structure of crystals of ErbB2 co-complexed with a variety of molecules, such as other EGF receptor family receptors to which ErbB2 dimerises, or chemical entities. For example, this approach enables the determination of the optimal sites for the interaction between chemical entities, and the interaction of candidate ErbB2 agonists or antagonists.

All of the complexes referred to above may be studied using well-known X-ray diffraction techniques and may be refined versus 1.5-3.5 A resolution X-ray data to an R value of about 0.25 or less using computer software, such as X-PLOR (Yale University, distributed by Molecular Simulations, Inc.; see Blundell & Johnson, supra; Meth. Enzymol., vol. 114 & 115, H. W. Wyckoff et al., eds., Academic Press (1985)). This information may thus be used to optimize known ErbB2 agonist/antagonists, such as anti-ErbB2 antibodies, and more importantly, to design new or improved ErbB2 agonists/antagonists.

Production of ErbB2 crystals

The crystals of the present invention may be prepared by expressing a nucleotide sequence encoding ErbB2 or a variant thereof in a suitable host cell, and then

crystallising the purified protein(s). Preferably the ErbB2 polypeptide contains the extracellular domain (amino acids 1 to 632 of the mature human polypeptide or a truncated version thereof, preferably comprising amino acids 1 to 509, or the equivalent residues in other ErbB2 polypeptides) but lacks the transmembrane and intracellular domains. Preferred host cells are those that provide for reduced glycosylation of recombinant polypeptides, such as a glycosylation-defective mammalian cell line e.g. the Lec8 Chinese hamster cell line, a derivative of CHO-K1 fibroblasts (ATCC CRC:1737) (Stanley, 1989, Mol. Cell Biol. 9: 377-383).

ErbB2 polypeptides may also be produced as fusion proteins, for example to aid in extraction and purification. Examples of fusion protein partners include glutathione-S-transferase (GST), hexahistidine, GAL4 (DNA binding and/or transcriptional activation domains) and beta-galactosidase. It may also be convenient to include a proteolytic cleavage site between the fusion protein partner and the protein sequence of interest to allow removal of fusion protein sequences.

After expression, the proteins may be purified and/or concentrated, for example by immobilised metal affinity chromatography, ion-exchange chromatography, and/or gel filtration.

The protein(s) may be crystallised using techniques described herein. Usually, in a crystallisation process, a crystallisation buffer is prepared with a lower concentration of a precipitating agent necessary for crystal formation. For crystal formation, the concentration of the precipitating agent has to be increased, by addition of precipitating agent or by diffusion of the precipitating agent between the crystallisation buffer and a reservoir buffer. Diffusion may be achieved by known techniques such as the "hanging drop" or the "sitting drop" method. In these methods, a drop of crystallisation buffer containing the protein (s) is hanging above or sitting beside a much larger pool of reservoir buffer. Alternatively, the balancing of the precipitating agent can be achieved through a semi-permeable membrane that separates the crystallisation buffer and prevents dilution of the protein into the reservoir buffer.

We have found that the inclusion of about 15% PEG 1500 provides optimal crystallization conditions for the extracellular domain of human ErbB2.

Generating the crystal structure

Once the crystals have been obtained, the structure may be solved by known X-ray diffraction techniques. Many techniques use chemically modified crystals, such as those modified by heavy atom derivatization. In practice, a crystal is soaked in a solution containing heavy metal atom salts, or organometallic compounds, e. g., lead chloride, gold thiomalate, thimerosal or uranyl acetate, which can diffuse through the crystal and bind to the surface of the protein. The location(s) of the bound heavy metal atom(s) can then be determined by X-ray diffraction analysis of the soaked crystal. The patterns obtained on diffraction of a monochromatic beam of X-rays by the atoms (scattering centres) of the crystal can be solved by mathematical equations to give mathematical coordinates. The diffraction data are used to calculate an electron density map of the repeating unit of the crystal. The electron density maps are used to establish the positions of the individual atoms within the unit cell of the crystal (Blundel, T. L. and N. L. Johnson, Protein Crystallography, Academic Press (1976)).

Target binding sites for modulators of ErbB2

The three-dimensional structure of ErbB2 provided herein allows the identification of target binding sites for potential ErbB2 modulators.

Preferred target binding sites are those involved in heterodimerisation of ErbB2 with other members of the EGF receptor family, such as ErbB1, ErbB3 and/or ErbB4.

One preferred binding site involved in heterodimerisation is the CR1 dimerisation loop (residues 247-268) and adjacent residues (residues 244-246, 285-289). Other suitable binding sites include the N-terminal end of the CR1 domain (residues 200-203, 210-213, 216-218, 225-230), and the C-terminal end of the CR1 domain (residues 294-319).

In a further preferred embodiment, the binding site is the docking site on ErbB2 for the CR1 dimerisation loop of heterodimer partners. This docking site is located on ErbB2 between the L1, CR1 and L2 domains. Preferably, the docking site comprises the following ErbB2 residues: Gln 36, Gln 60, Arg 82, Thr 84, Gln 85, Phe 237, Thr 269, Phe 270, Gly 271, Ala 272, Tyr 282, Thr 285, Gly 288, Ser 289, Cys 290, Thr 291, Leu 292, Val 293, Cys 294, Pro 295 and Cys 310.

In yet another preferred embodiment, the target binding site is located on the L1 or L2 domains. Unlike the unligated structure of ErbB3 (Cho, H. S. & Leahy, D. J. Science 297, 1330–1333 (2002)) or the pseudo-unligated structure of EGFR (Ferguson et al., Molecular Cell, Vol. 11, 507–517, (2003)), the structure of ErbB2 exists in a conformation similar to that of the 2:2 ligand-receptor dimer. This is in large part maintained by the L1:L2 contact, as described in Garrett, et al., Molecular Cell, Vol. 11, 495-505. Thus a small molecule or antibody which binds to either the L1 or L2 domain or intercalates between them can modulate receptor dimer formation by either preventing the domains from binding to each other or by modifying the relative positions of the domains. Thus binding of a chemical entity to the L1 and/or L2 domain may cause the protein to adopt a conformation similar to that of its unligated relatives (EGFR or ErbB3) and thereby inhibit dimerisation. Alternatively, binding of a chemical entity to the L1 and/or L2 domain may cause modifications in the CR1 (dimerisation domain) as described in Garrett, et al., Molecular Cell, Vol. 11, 495-505 to inhibit receptor dimer formation. The relevant binding sites of the L1 or L2 domain consist of the atoms of either one of these domains that lie within about 4.5 Angstroms of the other domain.

Antibodies

The term "antibody" as used in this invention includes intact molecules as well as fragments thereof, such as Fab, F(ab')2, and Fv which are capable of binding the epitopic determinant. These antibody fragments retain some ability to selectively bind with its antigen or receptor and are defined as follows:

(1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain;

(2) Fab', the fragment of an antibody molecule can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule;

(3) (Fab')2, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; F(ab)2 is a dimer of two Fab' fragments held together by two disulfide bonds;

(4) Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and

(5) Single chain antibody ("SCA"), defined as a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

Methods of making these fragments are known in the art. (See for example, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York (1988), incorporated herein by reference).

Antibodies of the present invention may be produced, for example, by immunizing mice with purified ErbB2 fragment 1-509. After determining that the mice are producing anti-ErbB2 antibodies, hybridomas may be prepared and antibody specificity assayed by ELISA or Flow Cytometry using two cell lines: Baf/wt-EGFR cells and Baf/EGFR-"mutation x" cells. These mouse cell lines express either the wild type ErbB2 or the ErbB2 containing an amino acid substitution, for example an Ala substitution (ie mutation x), within the specific site against which the antibody is to be directed. When hybridomas secreting antibodies which recognize Baf/wt-ErbB2, but not Baf/ErbB2-"mutant x" are identified, the corresponding hybridoma may be cloned and the monoclonal antibody purified.

Alternatively, in raising antibodies of the invention, it may be desirable to use derivatives of the peptides or loop structures which are conformationally constrained. Conformational constraint refers to the stability and preferred conformation of the three-dimensional shape assumed by a peptide. Conformational constraints include local constraints, involving restricting the conformational mobility of a single residue in a peptide; regional constraints, involving restricting the conformational mobility of a group of residues, which residues may form some secondary structural unit; and global constraints, involving the entire peptide structure. For example, amino acids adjacent to or flanking the ErbB2 loop structures may be included in the construct to maintain conformation of the peptide used to raise antibodies.

In addition, the active conformation of the peptide may be stabilized by a covalent modification, such as cyclization or by incorporation of gamma-lactam or other types of bridges. For example, side chains can be cyclized to the backbone so as create a L-gamma-lactam moiety on each side of the interaction site. See, generally, Hruby et al., "Applications of Synthetic Peptides," in Synthetic Peptides: A User's Guide: 259-345 (W. H. Freeman & Co. 1992). Cyclization also can be achieved, for example, by formation of cystine bridges, coupling of amino and carboxy terminal groups of respective terminal amino acids, or coupling of the amino group of a Lys residue or a related homolog with a carboxy group of Asp, Glu or a related homolog. Coupling of the alpha-amino group of a polypeptide with the epsilon-amino group of a lysine residue, using iodoacetic anhydride, can be also undertaken. See Wood and Wetzel, 1992, Int'l J. Peptide Protein Res. 39: 533-39.

Further the conformation of the peptide analogues may be stabilised by including amino acids modified at the alpha carbon atom (eg. α-amino-150-butyric acid) (Burgess and Leach, 1973, Biopolymers 12(12):2691-2712; Burgess and Leach, 1973, Biopolymers 12(11):2599-2605) or amino acids which lead to modifications on the peptide nitrogen atom (eg. sarcosine or N-methylalanine) (O'Donohue et al, 1995, Protein Sci. 4(10):2191-2202).

Another approach described in US 5,891,418 is to include a metal-ion complexing backbone in the peptide structure. Typically, the preferred metal-peptide backbone is based on the requisite number of particular coordinating groups required by the coordination sphere of a given complexing metal ion. In general, most of the metal ions that may prove useful have a coordination number of four to six. The nature of the coordinating groups in the peptide chain includes nitrogen atoms with amine, amide, imidazole, or guanidino functionalities; sulfur atoms of thiols or disulfides; and oxygen atoms of hydroxy, phenolic, carbonyl, or carboxyl functionalities. In addition, the peptide chain or individual amino acids can be chemically altered to include a coordinating group, such as for example oxime, hydrazino, sulfhydryl, phosphate, cyano, pyridino, piperidino, or morpholino. The peptide construct can be either linear or cyclic, however a linear construct is typically preferred.

Peptides and Peptidomimetics

In yet a further aspect the present invention provides an isolated conformationally constrained peptide or peptidomimetic consisting essentially of (i) ErbB2 amino acid residues 200-203, (ii) ErbB2 amino acid residues 210-213, (iii) ErbB2 amino acid residues 216-218, (iv) ErbB2 amino acid residues 225-230, (v) ErbB2 amino acid residues 247-268 or a subset thereof; (vi) ErbB2 amino acid residues 244-246, (vii) ErbB2 amino acid residues 285-289, or (viii) ErbB2 amino acid residues 294-319 or a subset thereof.

The term "conformationally constrained molecules" means conformationally constrained peptides and conformationally constrained peptide analogues and derivatives.

The term "analogues" refers to molecules having a chemically analogous structure to the naturally occurring alpha-amino acids present in ErbB2. Examples include molecules containing *gem*-diaminoalkyl groups or alklylmalonyl groups.

The term "derivatives" includes alpha amino acids wherein one or more side groups found in the naturally occurring alpha-amino acids present in ErbB2 have been modified. Thus, for example the naturally-occurring amino acids present in ErbB2 may be replaced with a variety of uncoded or modified amino acids such as the corresponding D-amino acid or N-methyl amino acid. Other modifications include substitution of hydroxyl, thiol, amino and carboxyl functional groups with chemically similar groups.

The present invention encompasses the use of conformationally constrained peptidomimetics of fragments of ErbB2 (such as amino acid residues 247-268), i.e. analogues and derivatives which mimic the activity of ErbB2 and are therefore capable of modulating ErbB2 activity *in vivo*. These peptidomimetics are preferably substantially similar in three-dimensional shape to the peptide structures (for example, loop structures) as they exist on the native ErbB2. Substantial similarity means that the geometric relationship of groups in the ErbB2 peptide fragment is preserved such that the peptidomimetic will mimic the activity of ErbB2 *in vivo*.

A "peptidomimetic" is a molecule that mimics the biological activity of a peptide but is no longer peptidic in chemical nature. By strict definition, a peptidomimetic is a molecule that no longer contains any peptide bonds (that is, amide bonds between amino acids). However, the term peptide mimetic is sometimes used to describe molecules that are no longer completely peptidic in nature, such as pseudo-peptides, semi-peptides and peptoids. Whether completely or partially non-peptide, peptidomimetics for use in the methods of the invention provide a spatial arrangement of reactive chemical moieties that closely resembles the three-dimensional arrangement of active groups in the peptide on which the peptidomimetic is based. As a result of this similar active-site geometry, the peptidomimetic has effects on biological systems which are similar to the biological activity of the peptide.

There are clear advantages for using a mimetic of a given peptide rather than the peptide itself, because peptides commonly exhibit two undesirable properties: (1) poor bioavailability; and (2) short duration of action. Peptide mimetics offer an obvious route around these two major obstacles, since the molecules concerned are small enough to be both orally active and have a long duration of action. There are also considerable cost savings and improved patient compliance associated with peptide mimetics, since they can be administered orally compared with parenteral administration for peptides. Furthermore, peptide mimetics are much cheaper to produce than peptides.

Suitable peptidomimetics based on, for example, residues 247-268, can be developed using readily available techniques. Thus, for example, peptide bonds can be replaced by non-peptide bonds that allow the peptidomimetic to adopt a similar structure, and therefore biological activity, to the original peptide. Further modifications can also be made by replacing chemical groups of the amino acids with other chemical groups of similar structure. The development of peptidomimetics derived from ErbB2 peptides based on residues 247-268 can be aided by reference to the three dimensional structure of these residues as provided in Appendix I. This structural information can be used to search three-dimensional databases to identify molecules having a similar structure, using programs such as MACCS-3D and ISIS/3D (Molecular Design Ltd., San Leandro, CA), ChemDBS-3D (Chemical Design Ltd., Oxford, U.K.), and Sybyl/3DB Unity (Tripos Associates, St. Louis, MO).

Those skilled in the art will recognize that the design of a peptidomimetic may require slight structural alteration or adjustment of a chemical structure designed or identified using the methods of the invention. In general, chemical compounds identified or designed using the methods of the invention can be synthesized chemically and then tested for ability to modulate ErbB2 activity using any of the methods described herein. The methods of the invention are particularly useful because they can be used to greatly decrease the number potential mimetics which must be screened ability to modulate ErbB2 activity.

The peptides or peptidomimetics of the present invention can be used in assays to screening for candidate compounds which bind to regions of ErbB2 and potentially interfere with the hereodimerisation of ErbB2 with another member of the EGF receptor family.

Standard solid-phase ELISA assay formats are particularly useful for identifying inhibitors of dimerisation. In accordance with this embodiment, the peptide or peptidomimetic of the invention is immobilized on a solid matrix, such as, for example an array of polymeric pins or a glass support. Conveniently, the immobilized peptide or peptidomimetic is a fusion polypeptide comprising Glutathione-S-transferase (GST; e.g. a CAP-ERK fusion), wherein the GST moiety facilitates immobilization of the protein to the solid phase support. This assay format can then be used to screen for candidate compounds that bind to the immobilised peptide or peptidomimetic and/or interefere with binding of a natural binding partner of ErbB2 to the immobilised peptide or peptidomimetic.

Uses of modulators of ErbB2

Compounds/chemical entities designed or selected by the methods of the invention described above may be used to modulate ErbB2 activity in cells, i.e. activate or inhibit ErbB2 activity. In particular, they may be used to modulate the interaction between ErbB2 and other heterodimerisation partners of the EGF receptor family, such as ErbB1, ErbB2 and ErbB4.

Modulation of heterodimerisation between ErbB2 and other members of the EGF receptor family may be achieved by direct binding of the chemical entity to a

heterodimerisation surface of ErbB2 and/or by an allosteric interaction elsewhere in the ErbB2 extracellular domain.

Given that aberrant EGF/ErbB2 activity is implicated in a range of disorders, the compounds described above may also be used to treat, ameliorate or prevent disorders characterised by abnormal ErbB2 signalling. Examples of such disorders include malignant conditions including tumours of the brain, head and neck, prostate, ovary, breast, cervix, lung, pancreas and colon; and melanoma, rhabdomyosarcoma, mesothelioma, squamous carcinomas of the skin and glioblastoma.

Administration

Compounds of the invention, i.e. antibodies of the invention or modulators of ErbB2 identified or identifiable by the screening methods of the invention, may preferably be combined with various components to produce compositions of the invention. Preferably the compositions are combined with a pharmaceutically acceptable carrier or diluent to produce a pharmaceutical composition (which may be for human or animal use).

The formulation will depend upon the nature of the compound and the route of administration but typically they can be formulated for topical, parenteral, intramuscular, oral, intravenous, intra-peritoneal, intranasal inhalation, lung inhalation, intradermal or intra-articular administration. The compound may be used in an injectable form. It may therefore be mixed with any vehicle which is pharmaceutically acceptable for an injectable formulation, preferably for a direct injection at the site to be treated, although it may be administered systemically.

The pharmaceutically acceptable carrier or diluent may be, for example, sterile isotonic saline solutions, or other isotonic solutions such as phosphate-buffered saline. The compounds of the present invention may be admixed with any suitable binder(s), lubricant(s), suspending agent(s), coating agent(s), solubilising agent(s). It is also preferred to formulate the compound in an orally active form.

In general, a therapeutically effective daily oral or intravenous dose of the compounds of the invention, including compounds of the invention and their salts, is likely to range from 0.01 to 50 mg/kg body weight of the subject to be treated, preferably 0.1 to 20

mg/kg. The compounds of the invention and their salts may also be administered by intravenous infusion, at a dose which is likely to range from 0.001-10 mg/kg/hr.

Tablets or capsules of the compounds may be administered singly or two or more at a time, as appropriate. It is also possible to administer the compounds in sustained release formulations.

Typically, the physician will determine the actual dosage which will be most suitable for an individual patient and it will vary with the age, weight and response of the particular patient. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

For some applications, preferably the compositions are administered orally in the form of tablets containing excipients such as starch or lactose, or in capsules or ovules either alone or in admixture with excipients, or in the form of elixirs, solutions or suspensions containing flavouring or colouring agents.

The compositions (as well as the compounds alone) can also be injected parenterally, for example intravenously, intramuscularly or subcutaneously. In this case, the compositions will comprise a suitable carrier or diluent.

For parenteral administration, the compositions are best used in the form of a sterile aqueous solution which may contain other substances, for example enough salts or monosaccharides to make the solution isotonic with blood.

For buccal or sublingual administration the compositions may be administered in the form of tablets or lozenges which can be formulated in a conventional manner.

For oral, parenteral, buccal and sublingual administration to subjects (such as patients), the daily dosage level of the compounds of the present invention and their pharmaceutically acceptable salts and solvates may typically be from 10 to 500 mg (in single or divided doses). Thus, and by way of example, tablets or capsules may contain from 5 to 100 mg of active compound for administration singly, or two or more at a time, as appropriate. As indicated above, the physician will determine the actual

dosage which will be most suitable for an individual patient and it will vary with the age, weight and response of the particular patient.

The routes of administration and dosages described are intended only as a guide since a skilled practitioner will be able to determine readily the optimum route of administration and dosage for any particular patient depending on, for example, the age, weight and condition of the patient.

The present invention will now be described further with reference to the following examples, which are illustrative only and non-limiting. The examples refer to the figures:

EXAMPLES

Experimental Procedures

Construction of the ErbB2-509 expression vector

An ErbB2 cDNA clone encompassing the entire coding region in the expression vector pRc/CMV (Invitrogen) was a gift from Dr. Rod Fiddes (AMBRI Pty.Ltd.). A Hind lll/EcoR 1 fragment spanning the 5'-non-coding region and nucleotides encoding amino acids 1-412 was isolated and cloned into pUC19 (Pharmacia). A 324 basepair EcoRl fragment incorporating amino acids 413-509 of ErbB2 and a C-terminal FLAG epitope (Brizzard et al., 1994, Biotechniques 16, 730-735) was generated by the polymerase chain reaction (PCR) using the primers 5'-CGGACAGCCTGCCTGACCTC-3' (upstream) and 5'-CCGGAATTCTAGACTACTTATCATCGTCATCTTTGTAATCGTTGACACA CTGGGTGGGC-3', and cloned into the EcoR 1 site of this plasmid. This plasmid was further modified by replacement of the 5' Hind III/BamH I of ErbB2 with a truncated Hind III/BamH I fragment, corresponding to nucleotides 171-1170 (GenBank accession number X03363), generated by PCR using the primers 5'-GGGGAAGCTTGCCACCATGGAGCTGGCGGCC-3' (upstream) and 5'-GCTGCACTTCTCACACCGCTG-3' (downstream). The fidelity of all amplification products was established by nucleotide sequencing. The modified ErbB2 cDNA insert was subsequently excised as a Hind III/Xba I fragment and cloned into the corresponding restriction sites of the mammalian expression vector pEE14 (Bebbington

and Hentschel, 1987, .In: DNA Cloning (Glover, D., ed.), Vol. III, pp.163-188, IRL Press, Oxford, U.K) to generate pESE.ErbB2-509.

Cell culture and transfection

The Lec8 Chinese hamster cell line, a derivative of CHO-K1 fibroblasts was obtained from the American Tissue Culture Collection (ATCC CRC:1737) and maintained in Glasgow's modified Eagle's medium (Life Technologies) supplemented with 10% fetal calf serum (FCS). Cells were transfected with pESE.ErbB2-509 that had been linearised by digestion with Fsp I, using FuGENE (Roche Molecular Biochemicals) according to the manufacturer's instructions. Stable transfectants were isolated by culturing cells in glutamine-free medium containing 10% dialysed FCS and 25 μM methionine sulfoximine. Supernatants were screened by dot-blotting onto nitrocellulose and probing with the anti-FLAG monoclonal antibody, M2 (Brizzard et al., 1994).

Protein Production and Purification

A positive polyclonal culture was used for scale-up protein production by growing the cells in roller bottles, during which time they were adapted to DMEM/F12 (JRH) media, supplemented with 10% dialysed FCS (Life Techologies) and 25uM methionine sulfoximine. After verifying the yield and quality of the ErbB2-509 fragment, four 500ml spinner flasks, each containing 10 g of FibraCell disks (New Brunswick Scientific), were inoculated with harvested cells from eight confluent roller bottles. Over a period of three weeks, spent media was collected daily from the spinner flasks and replaced with fresh media. Undialysed serum (CSL) was used instead of dialysed serum after day three. Approximately 30 litres of media harvest was collected over three weeks.

ErbB2-509 FLAG-tagged protein was purified by immunoaffinity chromatography over a 50 ml column of M2 anti-FLAG antibody covalently coupled to Mini Leak Low (Kem-En-Tek Denmark) as per manufacturer's instructions. Batches of four to six litres of culture media at 4 °C were passed over the column at 100 – 200 ml/h and washed with ~20 column volumes of 40 mM Tris-buffered saline at pH8 /0.02% sodium azide (TBSA). FLAG-tagged protein was eluted from the column after 90 min of recirculating 50 ml of a 0.25 mg/ml solution of the FLAG peptide DYKDDDDK in TBSA, followed by elution with three to four column volumes of 0.1 mg/ml FLAG

peptide in TBSA. The affinity column was regenerated with 0.1M sodium citrate pH 3 before re-equilibration at pH 8 with TBSA, ready for the next batch of harvest. Further purification was effected by passing a concentrated solution of the peptide-eluted product over a Superdex 200 column (Pharmacia 26/60) in TBSA at 5 ml/min. Greater than 90% of the 280nm-absorbing material eluted as a single symmetrical peak of apparent mass ~70 kDa , at a yield of 1-2mg/L of the spinner-flask harvest. The peak fraction was buffer-exchanged into 10 mM HEPES pH7.5 and concentrated to 8mg/ml.

Crystallization and Data Collection

Crystallization trials were performed with a factorial screen (Jancarik and Kim, 1991, J. Appl. Cryst. 24, 409-411) using the hanging drop method. Initially, rod-shaped crystals grew within 4 days, which diffracted to ~3.5 Å. However, after further crystallization trials the best conditions were 15% PEG 1500 and the resolution extended to 2.5 Å. Crystals (space group $P2_12_12_1$, a=75.96, b=82.24, c=110.06 Å) were cryo-cooled to –170 °C in 20% PEG, 20% glycerol. Diffraction data were recorded as 192 1° exposures on a Rigaku RAXIS IV area detector using RU-300 Rigaku generator equipped with elliptical glass capillary optics (AXCO). Data were integrated to 2.5 Å and scaled using the DENZO/SCALEPACK (mosaic spread 0.8°, R_{sym}=0.103, multiplicity 7.2, completeness 97.2%)

Structure Solution and Refinement

The structure was solved by molecular replacement with AMORE using data 10-4 Å resolution and two fragments of EGFR (residues 4-238 and 310-500) as search models. In both rotation and translation functions the highest peaks corresponded to the correct solution. By inspection of electron density maps (10-3.5 Å resolution) with O an initial model of ErbB2 was constructed from the structure of EGFR. This model consisted of 472 of 510 residues, including 91 side chains truncated from the EGFR equivalent. Structure refinement was performed with CNS (Brunger et al., 1998, Acta Crystallogr. D Biol Crystallogr. 54, 905-921). Initially, rigid body refinement with four groups (residues 1-194, 197-310, 318-510) gave R=0.473, R_{free}=0.482 (5% of the data). Nine rounds of manual refitting were alternated with energy minimisation, B factor refinement and, sometimes, simulated annealing. The resolution was extended in a stepwise manner with a bulk solvent correction applied from round 3 and an overall anisotropic thermal parameter from round 6. The final model contains 506 amino

acids, 4 carbohydrate residues and 134 solvent molecules, giving R=0.226, R_{free}=0.264 (data 25-2.5 Å). For residues 1-2, 100-102 and 107-113 294-318 the electron density is weak and there is no density for residues 103-106 or beyond residue 510.

Database Preparation

Databases were generated using information provided by the Supplier, or the NIH developmental therapeutics program. The NCI database was built from the October 2000 release, and the Tripos Leadquest database using the October 2001 release. SDF records were converted into 3-dimensional Sybyl mol2 files using the dbtranslate utility from UNITY environment in sybyl6.7, coordinates were generated using Concord 4.0.2 and the atom typing of resulting mol2 files corrected using our in house tool Mol-prepare. The resulting mol2 files were then protonated, assigned Gasteiger-Huckel charges and minimized (conjugate gradient for a maximum of 500 iterations) using Sybyl 6.7. Databases were then indexed for our database server program.

Assay for determining ErbB2 kinase activity

BaF/3 cells co-expressing K721R-ErbB1 and wtErbB2 are routinely grown in RPMI/10%FCS containing IL-3. Before assay, cells are washed three times to remove residual IL-3 and resuspended in RPMI 1640 + 10% FCS. Cells are seeded into 96 well plates using a Biomek 2000 (Beckman) at $2x10^4$ cells per 200µl and incubated for 4 hours at 37°C in 10% CO_2. Candidate ErbB2 inhibitors are added to the first titration point and titrated in two-fold dilutions across the 96 well plate in duplicate with or without a constant amount of mEGF (1nM) or IL-3 (1µl). ^{3H}-Thymidine (0.5µCi/well) is added and the plates incubated for 20 hours at 37°C in 5% CO_2. At the end of the incubation the cells are lysed in 0.5M NaOH at room temperature for 30 minutes then harvested onto nitrocellulose filter mats using an automatic harvester (Tomtec, Connecticut, USA). The mats are dried, placed in a plastic counting bag and scintillant (10ml) added. Incorporated 3H-Thymidine is determined using a beta counter (1205 Betaplate, Wallac, Finland).

Example 1: Description of structure

The ErbB2 fragment described here comprises the L1, CR1 and L2 domains plus the first module (residues 489-509) from the second cys-rich region CR2. The crystals contained only one molecule of the truncated ErbB2 ectodomain in the asymmetric unit and showed no evidence of dimers. ErbB2 (residues 1-509) adopts a compact bilobed structure reminiscent of the closed conformation of the EGFR ectodomain in its 2:2 complexes with TGFα (Garrett et al., 2002) or EGF (Ogiso et al., 2002, Cell 110, 775-787), but very different from the open conformations seen in the unliganded, full length ErbB3 ectodomain (Cho and Leahy, 2002) or the truncated L1/CR/L2 fragment of the closely related type 1 insulin-like growth factor receptor (Garrett et al., 1998, Nature 394, 395-399).

The main chain conformation of each L domain is similar to the corresponding domains of EGFR with the rmsd for the Cα atoms of L1 being 1.14-1.21 (for >91% of the Cα atoms) and for the Cα atoms of L2 being 0.97-1.05 Å (96%). In the ErbB2 L1 domain, the V-shaped region (residues 9-17), which forms a substantial part of the ligand-binding surface in EGFR, is maintained. However there is a small shift in the position of the N-terminal helix (residues 17-30) in ErbB2 and minor differences in residues equivalent to those in EGFR that make a main chain contact with TGFα (Garrett et al., 2002). The position of the large insertion (residues 101-109) specific to ErbB2 (Figure 1) is in the loop of EGFR (residues 101-106) in the fourth repeat at the corner of the second and third β-sheets of the L1 domain and is predominantly disordered in ErbB2. In the ErbB2 L2 main chain, small movements are seen in the two loops (residues 324-334 and 360-374) equivalent to those that bind ligand in EGFR (Garrett et al., 2002) and in the relative position of the single cys rich module (residues 489-509) that follows the L2 domain.

While the folds of the two L domains are similar in ErbB2 and the EGFR/TGFα complex (Garrett et al., 2002) the relative orientation of these two domains are quite different (Figures 3 and 4). This is due to differences in the CR1 domain and the CR1-L2 hinge of ErbB2 which direct the two L domains towards each other, where they make substantial contact (the total accessible surface area buried is 1264 $Å^2$ and shape complimentarily, S_c = 0.63). The overall movement of the ErbB2 L2 domain, with respect to L1, corresponds to a rotation of about 35 ° (A 37.4 °, B 31.8 °) around an axis parallel to strands of the L2 large β-sheet and a translation of 7 Å towards CR1 so that

in ErbB2 the bottom of the large sheet on L2 sits against the N-terminal end (residues 1-33) of L1. In this conformation an EGF-like ligand cannot bind to sites on either the L1 or L2 domains of ErbB2 (as seen for EGFR) since each site is occluded by the opposing L domain.

Finding ErbB2 in this "closed" form raises the question of whether this could occur in the unligated form of EGFR (or ErbB3/ErbB4). Structural superposition is not straightforward as there are main chain rearrangements in this region of ErbB2, namely a shift in the N-terminal helix of L1 by about 1.8 Å, possibly due to Leu22 of ErbB2 being replaced by an aromatic side chain in ErbB1 (F24), ErbB3 (Y27) and ErbB4 (Y24) (Figure 1). However, even when superimposing residues 9-17 and the helix separately then comparing with superpositions of L2, complementarity in these regions for other members of the family was not observed. In EGFR, Gln411 is equivalent to Ala419 of ErbB2 and the bulky Gln side chain could not be easily accommodated in the ErbB2 structure as it would sterically clash with Ser26 and Met30 (Met24 and Leu28 in ErbB2). This closed conformation would not pose a problem for ErbB3/ErbB4 where the residues corresponding to ErbB2 Ala419 are Gly residues (Figure 1). Asn12 in EGFR is a key residue for ligand binding and is strictly conserved in all the EGFR family except ErbB2. If the unliganded form of EGFR were to have the same conformation as ErbB2 then Asn12 (equivalent to Met10 in ErbB2) would sterically clash with His409 (Asn417 in ErbB2) and the side chains of Lys463 and Lys465 (equivalent to the hydrophobic residues Ala471 and Leu473 of ErbB2) on the last strand of the major β-sheet of L2, would overlap with Arg29 and Asp22, respectively. In addition to the steric clash, electrostatic repulsion may also be important as residues 29 and 463 are basic (Lys/Arg) in EGFR/ErbB3/Erb4 but are His and Ala in ErbB2, respectively. Thus it appears that the "closed" conformation seen for domains 1-3 of ErbB2 is unlikely to be a general feature of this receptor family but unique to the ErbB2 molecule.

Example 2: Analysis of the ligand binding region

A comparison of the ErbB2 and EGFR structures shows why ErbB2 does not bind ligands such as EGF, TGFα or the neuregulins. The L domains of EGFR, together with the ligand, TGFα, were superimposed on the corresponding L domains of ErbB2 using the strands of the large β-sheet. Residues of ErbB2 L1 which would interfere with ligand binding are Arg13 (replacing Thr, Ser and Ser in ErbB1, ErbB3 and ErbB4) and

Pro15 (replacing Leu, Thr and Leu in ErbB1, ErbB3 and ErbB4) on the short N-terminal strand of the L1 domain (Figure 1). In EGFR, residues 13-15 form a β-sheet with the ligand. The presence of Arg13 alone is likely to prevent ligand binding as this residue lies at the heart of the interface. Unless the receptor side chains are small there is no room for ligand side chains. Another crucial residue in EGFR is Asn12, the side chain of which makes two hydrogen bonds to the ligand's main chain. Asn is present in EGFR, ErbB3 and ErbB4 but in ErbB2 the equivalent residue (Met10) is buried between Val8 and Pro15 and unavailable for ligand interactions. Another residue in the L1 domain which would interfere with EGF-related ligand binding by ErbB2 is Asp98, which is Ser or Leu in the other ErbB family members (Figure 1) and would clash with Glu27 of TGFα

Observations by Kohda et al., 1993 (J. Biol. Chem. 268, 1976-1981) indicate that ligands can bind to L2 alone albeit with low affinity. For the L2 domain the differences between ErbB2 and the other ErbB receptors are more subtle. Asp355 of EGFR, which makes a salt bridge with the highly conserved Arg42 of TGFα (Garrett et al., 2002), is conserved for all EGFR homologues including ErbB2. However, in ErbB2 movement of residues 324-334 in a neighbouring loop appears to disturb the position of this residue (Asp363). Other residues in the L2 binding site of EGFR such as His346 and Gln384 are smaller in ErbB2 (Ala354 and Ser392), so binding to ErbB2 would be expected to be of lower affinity.

The ligand-binding surfaces of the EGFR homologues are by no means well conserved and each ErbB receptor has its own ligand binding characteristics. ErbB3 and ErbB4 predominantly bind the neuregulin group. Again, ErbB2 fails to interact with this subfamily of ligands and the residues of ErbB2 at positions equivalent to the EGFR ligand binding surface clearly disrupt the L1 and L2 binding surface (Figure 1).

Example 3: Differences in CR1

In the TGFα:EGFR complex, the dominant feature of CR1 is a large loop (residues 242-259) which extends out from the rod-shaped CR1 and plays a key role in homo-dimerisation and signaling for that receptor. This loop contains only limited sequence homology with the other EGFR homologues (33-44%) and it was not clear whether dimerisation of the receptor influenced the conformation of this loop. In the crystal, ErbB2 is present as a monomer and the CR1 loop projects out into solvent, lying

against an adjacent molecule in a crystal contact. Superposition of this loop from ErbB2 and EGFR shows that the main chain of the ErbB2 loop adopts a very similar structure to that found in the EGFR dimer (rmsd 0.61 A for 15 Cα's) with small differences seen only at the tip. Thus it seems that this loop has a well defined conformation even in the undimerised state. Residues important in maintaining this structure are prolines found at various positions in different homologues, particularly Pro257 (EGFR), and two completely conserved asparagines (residues 247 and 256 in EGFR) which make hydrogen-bonded contacts with main chain atoms. In ErbB2 this loop is bent slightly (12-13°) relative to the corresponding fifth cys rich module in EGFR.

Overall comparison of the CR1 domain shows that, relative to EGFR, it does not bend smoothly, rather it bends locally at three places. For individual modules rms deviations in Cα positions are less than 1 Å while for the whole domain the rms deviation is 1.7/1.8 Å (106/98 of 117 residues). Cys rich modules 2-4 lie similarly against L1 in both ErbB2 and EGFR and the bends occur at the interfaces of the fourth and seventh modules. The most obvious bend is in the middle of CR1 between the fourth and fifth modules (21/25°) but other differences (between the fifth and sixth modules 11°, between the sixth and seventh modules 15/27°), together with a bend of 37/32° between the seventh module and the L2 domain constitute the set of changes which reorientate L2 with respect to L1.

Example 4: Implications for dimerisation

The rearrangements in CR1 have three effects on the dimer interface as seen in EGFR and the capacity of ErbB2, in this conformation, to form heterodimers with a 1:1 complex of EGFR with ligand. Superposing the fifth cys-rich module from CR1 of ErbB2 on one half of the EGFR dimer, the bend at the fourth and fifth modules of ErbB2 causes the N-terminal tip of ErbB2 to move away from the corresponding region on the other molecule, removing that region from the back-to-back contact. The bend at cys-rich modules 6 and 7 of ErbB2 would bring module 8 in contact with module 7 of EGFR. More significant, however, is that the bend at the fourth and fifth modules of ErbB2 brings the ErbB2 L1domain closer to the tip of the partner's CR1 loop, causing Thr249 of EGFR) to overlap with Thr84 and Gln60 of ErbB2. Therefore it seems unlikely that ErbB2 could interact with EGFR in the closed form. With some minor

structural rearrangement the tip of ErbB2's CR1 loop could be accommodated in EGFR.

Example 5: *In silico* screening for compounds that modulate ErbB2 activity

Molecular docking of large compound data bases to target proteins of known or modeled 3-dimensional structure is now a common approach in the identification of new lead compounds. This "virtual screening" approach relies on fast and accurate estimation of the ligand binding mode and an estimate of ligand affinity. Typically a large database of compounds, either real or virtual is docked to a target structure and a list of the best potential ligands is produced. This ranking should be highly enriched for active compounds which may then be subject to further experimental validation.

The calculation of the ligand binding mode may carried out by molecular docking programs which are able to dock the ligands in a flexible manner to a static protein structure. The estimation of ligand affinity is typically carried out by the use of a separate scoring function. These scoring functions include empirical functions [DOCK potential energy, Chemscore, Score], or knowledge based potentials of mean force [PMF, SMoG]. Consensus scoring involves re-scoring each ligand with multiple scoring functions and then using a combination of these rankings to generate a hit list.

We used the program DOCK (vers. 4.0.1) for the generation of favorable conformations of ligand binding. Protons and Kollman all-atom charges were added to the protein using the Biopolymer module of Sybyl6.7 and proton positions minimized with all other atoms held fixed. Scoring grids were calculated using the GRID program with a grid resolution of 0.25 Angstrom. All conformations were minimized using the DOCK energy function. Docking of ligand databases were directed towards the sites identified previously. Nine scoring functions were used, including Score, the Score-Quality estimate, DOCK energy function, PMF, PMF-RB (the PMF function with penalties for rotatable bonds), the SMoG function, SMoG/H (the SMOG function scaled by the number of ligand heavy atoms), Chemscore, and the Autodock Scoring function. Ligand conformations were chosen using a rank-by-rank consensus of the nine different scoring methods of the best 25 solutions obtained from the DOCK program using the DOCK potential energy. A ranked list of compounds was generated using a consensus of the individual scores for each ligand (in their best consensus-ranked conformation).

Four of the top-ranked compounds (compounds 39293, 94289, 19378 and 20697) were obtained and tested for their ability to modulate ErbB2 kinase activity according to the method described above. The results of these inhibition assays are shown in Figure 3. These results show that all four compounds tested inhibited ErbB2 kinase activity at concentrations of between 10^{-1} and 10^{2} μM.

Conclusion

The availability of 3D structures for the ErbB2(1-509) monomer, the 2:2 dimer complexes of TGFα:sEGFR501 (Garrett et al., 2002) and EGF:EGFR621 (Ogiso et al., 2002), the unliganded ErbB3 ectodomain monomer (Cho and Leahy, 2002) and the related L1/cys-rich/L2 fragment of the IGF-1R (Garrett et al., 1998) provides the framework to explore some of the outstanding issues related to ErbB receptor function. A striking feature in these comparisons is the flexibility that exists at the CR1/L2 junction resulting in major differences in the positioning of the L2 domain relative to the L1/CR1 region.

The structure of the unliganded ErbB3 full length ectodomain is even more open than that of the IGF-1R fragment, with the L2 domain rotated further away from the L1 domain (Figure 3). This open conformation is very different from the closed arrangement of the L1 and L2 domains seen in the two EGFR/ligand dimer structures and in the ErbB2(1-509) structure reported here. The open conformation is stabilised by a single main chain/main chain hydrogen bond and side chain interactions between residues Tyr246, Phe251 and Gln252 in the CR1 loop (residues 242-259) and Asp562, Gly563, His565 (module 5) and Lys583 (module 6) of CR2. These contact residues are conserved in EGFR and ErbB4 but not in ErbB2 (Cho and Leahy, 2002). This open structure of ErbB3 provides an explanation for the predominantly low affinity ligand binding by soluble full length EGFR ectodomain compared to the high affinity binding shown by sEGFR501, which cannot make these contacts since it lacks CR2 modules 2 to 7. The same CR1 loop is critically involved in formation of the ligand-induced EGFR dimers suggesting that it becomes available for such dimer interactions following ligand binding.

The "closed" structure of the unliganded ErbB2(1-509) fragment seen here, where the bottom of the L2 domain sits against the top of the L1 domain resembles a "pseudo-

active" arrangement of domains, similar to that seen in the EGFR/ligand complexes (Garrett et al., 2002; Ogiso et al., 2002). It may represent the conformation of the full length ectodomain, since the residues involved in the ErbB3 CR1 loop/CR2 interactions are not conserved in ErbB2 (Cho and Leahy, 2002) and such a constraining CR1/CR2 interaction may not be tolerated in a receptor that does not bind ligand TGFα:sEGFR501.

The 3D structure of ErbB2 also allows the epitopes for monoclonal antibodies to be mapped and their mode of action inferred, since some inhibit, some stimulate and others have no effect on cell growth. The epitopes for mAbs L87, N28 and N12 have been located to the regions Cys199-Cys214, Thr195-Cys214 and Cys510-Ala565 (mature receptor numbering) respectively (Yip YL, Smith G, Koch J, Dubel S, Ward RL. Identification of epitope regions recognized by tumor inhibitory and stimulatory anti-ErbB-2 monoclonal antibodies: implications for vaccine design. J Immunol. 166(8):5271-8, (2001)). The epitopes for mAbs L87 and N28 (reported to have no effect or to stimulate growth of a subset of breast cancer cell lines respectively) are located in the second cys rich module of CR1, while the epitope for mAb N12, an inhibitory antibody, is located within a large region comprising cys rich modules 2 to 4 of CR2 (Figure 2). Similarly the epitope for the potential therapeutic anti-ErbB2 monoclonal antibody MGr6 (Orlandi R, Formantici C, Menard S, Boyer CM, Wiener JR, Colnaghi M. A linear region of a monoclonal antibody conformational epitope mapped on p185HER2 oncoprotein. J. Biol Chem. 378(11):1387-92, (1997)) has been shown to include residues 207-215 (mature receptor numbers) in the third module of CR1.

The CR2 region has also been implicated as the site of action for a set of inhibitory peptides originally designed to mimic the CDR3 loop of herceptin and shown to compete with herceptin for binding to ErbB2. A subsequent set of inhibitory peptides have been designed which mimic sequences in modules 4 to 6 of CR2, a region shown to contribute to ErbB2 heterodimer formation. Other inhibitors of ErbB2 function include the ErbB2 splice variant herstatin and the small, leucine-rich repeat proteoglycan decorin. The inhibition of ErbB2 function in breast cancer cells by decorin has been shown to be indirect and involves inactivation of ErbB4, presumably by direct binding.

The availability of the 3D structures of these receptors will facilitate the determination of the precise mechanism of action of these inhibitory agents and the design of new approaches to interfering with ErbB receptor function.

The disclosure of all publications referred to in this application are incorporated herein by reference.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.